An upper variance bound for the multinomial and the negative multinomial distribution [*]

G. Afendras[†] and V. Papathanasiou

Department of Mathematics, Section of Statistics and O.R., University of Athens, Panepistemiopolis, 157 84 Athens, Greece

Abstract

We prove a Chernoff-type upper variance bound for the multinomial and the negative multinomial distribution.

MSC: Primary 60E15.

Key words and phrases: Cauchy-Schwartz inequality; Covariance identity; Variance bounds

1 Introduction

Let X be a standard normal distribution and g be an absolutely continuous function, with a.s. derivative g'. Chernoff [14] proved that $\operatorname{Var} g(Z) \le \mathbb{E}\left(g'(Z)\right)^2$, provided that $\mathbb{E}\left(g'(Z)\right)^2$ is finite, where the equality holds if and only if g is a polynomial of degree at most one; see also the previous papers by Nash [20], Brascamp and Lieb [7]. This inequality has been generalized and extended by many authors (see, e.g., [13, 8, 9, 10, 19, 18, 22, 6, 16, 15, 17, 4, 5, 12, 23, 2, 3, 1, 24]).

Let X be an integer-valued random variable (r.v.) with finite mean μ, finite variance σ^2 and probability mass function (p.m.f.) p. And let the function w be defined by

$$\textstyle\sum_{j \le x} (\mu - j) p(j) = \sigma^2 w(x) p(x) \quad \text{for all } \; x \in \mathbb{Z}.$$

In case where w is a quadratic polynomial (of degree at most 2) and for any suitable function g, defined on the support of X, Cacoulos and Papathanasiou [9] proved that (see also Afendras et al. [4])

$$\operatorname{Var} g(X) \le \sigma^2 \mathbb{E} w(X) [\Delta g(X)]^2, \tag{1.1}$$

where Δ is the forward difference operator. Furthermore, the following Stein-type covariance identity holds (see Cacoullos and Papathanasiou [9], Afendras et al. [5])

$$\operatorname{Cov}[X, g(X)] = \sigma^2 \mathbb{E} w(X) \Delta g(X). \tag{1.2}$$

Cacoullos and Papathanasiou [11] extended this identity for discrete multivariate distributions, see Appendix A, and established a lower variance bound for the variance of $g(\boldsymbol{X})$,

[*]Work partially supported by the University of Athens' Research fund under Grant 70/4/5625

[†]Corresponding author. `e-mail address: g_afendras@math.uoa.gr`

where g is a suitable real function defined on the support of $\boldsymbol{X}$. For the multinomial and the negative multinomial distribution this bound takes the form

$$\operatorname{Var} g(\boldsymbol{X}) \geq \mathbb{E}\big(w(\boldsymbol{X})\nabla^{\mathrm{t}} g(\boldsymbol{X})\big)\, \boldsymbol{\Sigma}\, \mathbb{E}\big(w(\boldsymbol{X})\nabla g(\boldsymbol{X})\big),$$

where $\boldsymbol{\Sigma}$ is the dispersion matrix of $\boldsymbol{X}$, the function w is given by (2.3) for the multinomial case and by (2.5) for the negative multinomial case, and ∇g is the grad of g (see Definition 2.1).

Chen [13] extended Chernoff's bound to the multivariate normal case. Specifically, let $\boldsymbol{X}$ be the k-dimensional normal distribution $N_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Then,

$$\operatorname{Var} g(\boldsymbol{X}) \leq \mathbb{E}\big(\nabla^{\mathrm{t}} g(\boldsymbol{X})\, \boldsymbol{\Sigma}\, \nabla g(\boldsymbol{X})\big),$$

where $\nabla g(\boldsymbol{X}) = (\partial g(\boldsymbol{x})/\partial x_1, \ldots, \partial g(\boldsymbol{x})/\partial x_k)^{\mathrm{t}}$ is the grad of g (cf. Definition 2.1(a,b)).

In this note we extend (1.1) for multrinomial and negative multinomial distributions. Specifically, we prove that

$$\operatorname{Var} g(\boldsymbol{X}) \leq \mathbb{E}\big(w(\boldsymbol{X})\nabla^{\mathrm{t}} g(\boldsymbol{X})\, \boldsymbol{\Sigma}\, \nabla g(\boldsymbol{X})\big).$$

2 Preliminaries

The following definitions will be used in the sequel.

Definition 2.1 *Consider the vectors* $\boldsymbol{x} = (x_1, \ldots, x_k)^{\mathrm{t}} \in \mathbb{R}^k$ *and* $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_k)^{\mathrm{t}} \in (0,1)^k$, *a non-negative integer* ν *and a real function* g *defined on* $\mathbb{R}^k$. *We define:*

(a) $g_i(\boldsymbol{x}) \equiv \Delta_i g(\boldsymbol{x}) := g(\boldsymbol{x} + \boldsymbol{e}_i) - g(\boldsymbol{x})$, *where* $\boldsymbol{e}_i$ *is the* i*-th vector of the standard orthonormal basis of* $\mathbb{R}^k$;

(b) $\nabla^{\mathrm{t}} g(\boldsymbol{x}) \equiv (\nabla g(\boldsymbol{x}))^{\mathrm{t}} := (g_1(\boldsymbol{x}), g_2(\boldsymbol{x}), \ldots, g_k(\boldsymbol{x}))$.

(c) $\boldsymbol{\pi}^{\boldsymbol{x}} := \pi_1^{x_1} \cdots \pi_k^{x_k}$.

(d) $\binom{\nu}{\boldsymbol{x}} := \nu!/[x_1! \cdots x_k!(\nu - x_1 - \cdots - x_k)!]$, *provided that* $\boldsymbol{x} \in \mathbb{N}^k$ *with* $\sum_{i=1}^k x_i \leq \nu$.

(e) $\boldsymbol{x}_{-k} := (x_1, \ldots, x_{k-1})^{\mathrm{t}} \in \mathbb{R}^{k-1}$.

Definition 2.2 *Let* $\boldsymbol{X} = (X_1, \ldots, X_k)^{\mathrm{t}}$ *be a discrete random vector. We denote by:*

(a) $\mathrm{b}(n, \pi)$ *the binomial distribution with p.m.f.* $p(x) = \binom{n}{x}\pi^x(1-\pi)^{n-x}$, $x = 0, 1, \ldots, n$, *(*$n \in \mathbb{N}$*).*

(b) $\mathrm{nb}(r, \pi)$ *the negative binomial distribution with p.m.f.* $p(x) = \binom{r+x-1}{x}\pi^r(1-\pi)^x$, $x = 0, 1, \ldots$, *(*$r \in \mathbb{N}^* \equiv \mathbb{N} \setminus \{0\}$*).*

(c) $\mathrm{m}_k(n, \boldsymbol{\pi})$ *the* k*-dimensional multinomial distribution with p.m.f.* $p(\boldsymbol{x}) = \binom{n}{\boldsymbol{x}}\boldsymbol{\pi}^{\boldsymbol{x}}\pi_0^{x_0}$, $\boldsymbol{x} \in \mathbb{N}^k$ *with* $\sum_{i=1}^k x_i \leq n$, *where* $x_0 := n - \sum_{i=1}^k x_i$, $\boldsymbol{\pi} \in (0,1)^k$ *and* $\pi_0 := 1 - \sum_{i=1}^k \pi_i > 0$.

(d) $\mathrm{nm}_k(r, \boldsymbol{\theta})$ *the k-dimensional negative multinomial distribution with p.m.f.* $p(\boldsymbol{x}) = \binom{r+\sum_{i=1}^k x_i - 1}{\boldsymbol{x}} \boldsymbol{\theta}^{\boldsymbol{x}} \theta_0^r$, $\boldsymbol{x} \in \mathbb{N}^k$, *where* $\boldsymbol{\theta} \in (0,1)^k$ *and* $\theta_0 := 1 - \sum_{i=1}^k \theta_i > 0$.

(e) $p_k(\boldsymbol{x}) \equiv p_{X_k}(x_k)$ *and* $p_{-k}(\boldsymbol{x}) \equiv p_{\boldsymbol{X}_{-k}}(\boldsymbol{x}_{-k})$ *the p.m.f.s of marginal* X_k *and* $\boldsymbol{X}_{-k}$ *of* $\boldsymbol{X}$, *respectively, and* $p_{-k|k}(\boldsymbol{x}) \equiv p_{\boldsymbol{X}_{-k}|X_k=x_k}(\boldsymbol{x}_{-k})$ *the p.m.f. of the conditional r.v.* $\boldsymbol{X}_{-k}|X_k = x_k$.

Let $\boldsymbol{X} \sim \mathrm{m}_k(n, \boldsymbol{\pi})$. We define the function $w(\boldsymbol{x}) \equiv w_{\boldsymbol{X}}(\boldsymbol{x})$ by

$$w(\boldsymbol{x}) := \begin{cases} \frac{n - \sum_{i=1}^k x_i}{n\pi_0}, & \text{when } n \in \mathbb{N}^*, \\ 0, & \text{when } n = 0, \end{cases} \tag{2.3}$$

where π_0 as above. Notice that in the case $n = 0$ for each h the r.v. $h(\boldsymbol{X})$ is a constant with prob. 1 [$\mathrm{Var}\, h(\boldsymbol{X}) = 0$]; thus, we define $w = 0$. It is obvious that $X_k \sim \mathrm{b}(n, \pi_k)$ and $\boldsymbol{X}_{-k}|X_k = x_k \sim \mathrm{m}_{k-1}(n - x_k, \boldsymbol{\varpi})$, where $\boldsymbol{\varpi} \in (0,1)^{k-1}$ with $\varpi_i = \frac{\pi_i}{1-\pi_k}$, $i = 1, \ldots, k-1$. Thus, we define the functions $w_k(\boldsymbol{x}) \equiv w_{X_k}(x_k)$ and $w_{-k|k}(\boldsymbol{x}) \equiv w_{\boldsymbol{X}_{-k}|X_k=x_k}(\boldsymbol{x}_{-k})$ by

$$w_k(\boldsymbol{x}) := \begin{cases} \frac{n-x_k}{n(1-\pi_k)}, & \text{when } n \in \mathbb{N}^*, \\ 0, & \text{when } n = 0, \end{cases} \qquad w_{-k|k}(\boldsymbol{x}) := \begin{cases} \frac{(1-\pi_k)\left(n - \sum_{i=1}^k x_i\right)}{(n-x_k)\pi_0}, & \text{when } x_k < n, \\ 0, & \text{when } x_k = n, \end{cases} \tag{2.4}$$

Let $\boldsymbol{X} \sim \mathrm{nm}_k(r, \boldsymbol{\theta})$. Then $X_k \sim \mathrm{nb}(r, \vartheta_k)$, where $\vartheta_k = \frac{\theta_k}{\theta_0 + \theta_k}$, and $\boldsymbol{X}_{-k}|X_k = x_k \sim \mathrm{nm}_{k-1}(r + x_k, \boldsymbol{\theta}_{-k})$. Thus, similarly, we define the functions w, w_k and $w_{-k|k}$ by

$$w(\boldsymbol{x}) := \frac{\theta_0\left(r + \sum_{i=1}^k x_i\right)}{r}, \quad w_k(\boldsymbol{x}) := \frac{\theta_0(r + x_k)}{r(\theta_0 + \theta_k)} \quad \text{and} \quad w_{-k|k}(\boldsymbol{x}) := \frac{(\theta_0 + \theta_k)\left(r + \sum_{i=1}^k x_i\right)}{r + x_k}. \tag{2.5}$$

For both cases (multinomial and negative multinomial distribution) one can easily see that

$$p_{-k|k}(\boldsymbol{x} + \boldsymbol{e}_k) = p_{-k|k}(\boldsymbol{x}) w_{-k|k}(\boldsymbol{x}) \quad \text{and} \quad w_k(\boldsymbol{x}) w_{-k|k}(\boldsymbol{x}) = w(\boldsymbol{x}). \tag{2.6}$$

Next, we prove the following useful lemma.

Lemma 2.1 *Let* $\boldsymbol{X} \sim \mathrm{m}_k(n, \boldsymbol{\pi})$ *or* $\mathrm{nm}_k(r, \boldsymbol{\theta})$. *Consider a real function g defined on support of* $\boldsymbol{X}$ *such that* $\mathbb{E}|X_j g(\boldsymbol{X})|$ *and* $\mathbb{E}|X_j g_i(\boldsymbol{X})|$ *are finite for all* $i, j = 1, \ldots, k$.

(a) *The following covariance identity holds*

$$\mathrm{Cov}\left[\sum_{i=1}^k X_i, g(\boldsymbol{X})\right] = \mathbb{E}\left(w(\boldsymbol{X}) \sum_{i=1}^k c_i g_i(\boldsymbol{X})\right), \tag{2.7}$$

where w is given by (2.3) *or* (2.5), *respectively, and* $c_i = \sum_{j=1}^k \sigma_{ij}$ *with* $\sigma_{ij} = \mathrm{Cov}(X_i, X_j)$.

(b) *The next identity is valid (for multinomial only when* $X_k < n$*)*

$$\Delta_k \mathbb{E}[g(\boldsymbol{X})|X_k] = \mathbb{E}\left[w_{-k|k}(\boldsymbol{X})\left(g_k(\boldsymbol{X}) + \alpha_k \sum_{i=1}^{k-1} c_{i|k} g_i(\boldsymbol{X})\right)\Big| X_k\right], \tag{2.8}$$

where $c_{i|k} = \sum_{j=1}^{k-1} \sigma_{ij|k}$ *with* $\sigma_{ij|k} = \mathrm{Cov}(X_i, X_j|X_k)$ *and* $a_k \equiv a(X_k)$ *is* $-(1-\pi_k)/[\pi_0(n - X_k)]$ *for the multinomial and is* $(\theta_0 + \theta_k)/(r + X_k)$ *for the negative multinomial .*

Proof (a) For multinomial and negative multinomial distributions the identity (A.10) is valid. Note that for both cases $w^i(\boldsymbol{x}) = w(\boldsymbol{x})$, for each $i = 1, \ldots, k$, where the function $w(\boldsymbol{x})$ is given by (2.3) or (2.5), respectively (see Cacoullos and Papathanasiou [11, pp. 178–179]). So,

$$\mathrm{Cov}\,[q^i(\boldsymbol{X}), g(\boldsymbol{X})] = \mathbb{E}w(\boldsymbol{X})g_i(\boldsymbol{X}), \quad \text{for all} \;\; i = 1, \ldots, k.$$

Now, by (A.8) it follows that $\boldsymbol{X} = \boldsymbol{\Sigma}\, \boldsymbol{q}(\boldsymbol{X})$. Therefore, $\sum_{i=1}^{k} X_i = \sum_{i=1}^{k} c_i q^i(\boldsymbol{X})$. Combining the above relations (2.7) follows.

(b) We write $\Delta_k \mathbb{E}[g(\boldsymbol{X})|X_k] = \mathbb{E}[g(\boldsymbol{X})|X_k+1] - \mathbb{E}[g(\boldsymbol{X})|X_k]$. Using (2.6), $\Delta_k \mathbb{E}[g(\boldsymbol{X})|X_k] = \mathbb{E}\left[w_{-k|k}(\boldsymbol{X}) g_k(\boldsymbol{X})|X_k\right] + \mathbb{E}\left[w_{-k|k}(\boldsymbol{X}) g(\boldsymbol{X})|X_k\right] - \mathbb{E}[g(\boldsymbol{X})|X_k]$. Since $\mathbb{E}\left[w_{-k|k}(\boldsymbol{X})|X_k\right] = 1$, it follows that $\Delta_k \mathbb{E}[g(\boldsymbol{X})|X_k] = \mathbb{E}\left[w_{-k|k}(\boldsymbol{X}) g_k(\boldsymbol{X})|X_k\right] + \mathrm{Cov}\left[w_{-k|k}(\boldsymbol{X}), g(\boldsymbol{X})|X_k\right]$. From (2.3) and (2.5) we observe that $w_{-k|k}(\boldsymbol{X}) = \alpha_k \sum_{i=1}^{k-1} X_i + \beta_k$, where $\beta_k \equiv \beta(X_k)$ is a constant in $X_1, \ldots, X_{k-1}$. Thus,

$$\Delta_k \mathbb{E}[g(\boldsymbol{X})|X_k] = \mathbb{E}\left[w_{-k|k}(\boldsymbol{X}) g_k(\boldsymbol{X})|X_k\right] + \alpha_k \,\mathrm{Cov}\left[\, \textstyle\sum_{i=1}^{k-1} X_i, g(\boldsymbol{X})|X_k\right].$$

Finally, from the conditions on g it follows that $\mathbb{E}\big|X_j g(\boldsymbol{X})|X_k\big| < \infty$, for all $j = 1, \ldots, k-1$, and $\mathbb{E}\big|X_j g_i(\boldsymbol{X})|X_k\big| < \infty$ for all $i, j = 1, \ldots, k-1$. Thus, using (2.7) for $\boldsymbol{X}_{-k|k}$ the lemma is proved. □

3 The main result

We are now in a position to state and prove the main result.

Theorem 3.1 *Let $\boldsymbol{X} \sim \mathrm{m}_k(n, \boldsymbol{\pi})$ or $\mathrm{nm}_k(r, \boldsymbol{\theta})$. Consider a function g defined on support of $\boldsymbol{X}$; for the negative multinomial assume further that $\mathrm{Var}\, g(\boldsymbol{X})$ is finite. Then,*

$$\mathrm{Var}\, g(\boldsymbol{X}) \leq \mathbb{E}\left[w(\boldsymbol{X}) \nabla^{\mathrm{t}} g(\boldsymbol{X})\, \boldsymbol{\Sigma}\, \nabla g(\boldsymbol{X})\right], \tag{3.1}$$

where $\boldsymbol{\Sigma}$ is the dispersion matrix of $\boldsymbol{X}$ and the function w for the multinomial is given by (2.3) *and for the negative multinomial is given by* (2.5). *The equality in* (3.1) *holds if and only if g is a linear function with respect to $x_1, \ldots, x_k$, i.e. of the form $g(\boldsymbol{x}) = \rho_0 + \sum_{i=1}^{k} \rho_i x_i$.*

Proof If $\mathbb{E}\left[w(\boldsymbol{X}) \nabla^{\mathrm{t}} g(\boldsymbol{X})\, \boldsymbol{\Sigma}\, \nabla g(\boldsymbol{X})\right]$ is infinite the relation (3.1) is trivial. Assume that $\mathbb{E}\big[w(\boldsymbol{X}) \nabla^{\mathrm{t}} g(\boldsymbol{X})\, \boldsymbol{\Sigma}\, \nabla g(\boldsymbol{X})\big]$ is finite and thus the conditions of Lemma 2.1 are valid. The proof will be done by induction on k. For $k = 1$ (3.1) holds, see (1.1). Assuming that (3.1) is valid for $k-1$ for some $k > 1$, we will prove that (3.1) is also valid for k. It is well know that

$$\mathrm{Var}\, g(\boldsymbol{X}) = \mathbb{E}[\mathrm{Var}\,(g(\boldsymbol{X})|X_k)] + \mathrm{Var}\,[\mathbb{E}(g(\boldsymbol{X})|X_k)]\,. \tag{3.2}$$

Using (1.1) for X_k it follows that

$$\mathrm{Var}\,[\mathbb{E}(g(\boldsymbol{X})|X_k)] \leq \sigma_k^2 \mathbb{E}\big[w_k(\boldsymbol{X})\,(\Delta_k \mathbb{E}(g(\boldsymbol{X})|X_k))^2\,\big], \tag{3.3}$$

where $\sigma_k^2 = \mathrm{Var}\, X_k$. Note that $w_k(\boldsymbol{X})|_{X_k=n} = 0$. Thus, from (2.8) it follows that

$$\mathrm{Var}\,[\mathbb{E}(g(\boldsymbol{X})|X_k)] \le \sigma_k^2 \mathbb{E}\big[w_k(\boldsymbol{X})\big(\mathbb{E}\big[w_{-k|k}(\boldsymbol{X})\big(g_k(\boldsymbol{X}) + \alpha_k \textstyle\sum_{i=1}^{k-1} c_{i|k} g_i(\boldsymbol{X})\big)\big|X_k\big]\big)^2\big].$$

Since $\mathbb{E}[w_{-k|k}(\boldsymbol{X})|X_k] = 1$, using Cauchy–Schwartz inequality it follows that

$$\begin{aligned} &\mathbb{E}^2\big[w_{-k|k}(\boldsymbol{X})\big(g_k(\boldsymbol{X}) + \alpha_k \textstyle\sum_{i=1}^{k-1} c_{i|k} g_i(\boldsymbol{X})\big)\big|X_k\big] \\ &\qquad \le \mathbb{E}\Big[w_{-k|k}(\boldsymbol{X})\big(g_k(\boldsymbol{X}) + \alpha_k \textstyle\sum_{i=1}^{k-1} c_{i|k} g_i(\boldsymbol{X})\big)^2\big|X_k\Big]. \end{aligned} \tag{3.4}$$

Using (2.6),

$$\begin{aligned} &\mathbb{E}[\mathrm{Var}\,(g(\boldsymbol{X})|X_k)] \\ &\quad \le \mathbb{E}\mathbb{E}\big[\sigma_k^2 w(\boldsymbol{X})\big(g_k^2(\boldsymbol{X}) + 2\alpha_k \textstyle\sum_{i=1}^{k-1} c_{i|k} g_i(\boldsymbol{X}) g_k(\boldsymbol{X}) + \big(\alpha_k \sum_{i=1}^{k-1} c_{i|k} g_i(\boldsymbol{X})\big)^2\big)\big|X_k\big] \\ &\quad = \mathbb{E}\big[\sigma_k^2 w(\boldsymbol{X})\big(g_k^2(\boldsymbol{X}) + 2\alpha_k \textstyle\sum_{i=1}^{k-1} c_{i|k} g_i(\boldsymbol{X}) g_k(\boldsymbol{X}) + \alpha_k^2\big(\sum_{i=1}^{k-1} c_{i|k} g_i(\boldsymbol{X})\big)^2\big)\big]. \end{aligned} \tag{3.5}$$

By the inductional assumption (3.1) (with $k-1$ in place of k) it follows that

$$\mathrm{Var}\,(g(\boldsymbol{X})|X_k) \le \mathbb{E}\big[w_{-k|k}(\boldsymbol{X}) \nabla_{-k}^{\mathrm{t}} g(\boldsymbol{X})\, \boldsymbol{\Sigma}_{-k|k} \nabla_{-k} g(\boldsymbol{X})\big|X_k\big], \tag{3.6}$$

where $\boldsymbol{\Sigma}_{-k|k}$ is the dispersion matrix of $\boldsymbol{X}_{-k|k}$ and $\nabla_{-k} g = (g_1, \ldots, g_{k-1})^{\mathrm{t}}$. Thus,

$$\begin{aligned} \mathbb{E}[\mathrm{Var}\,(g(\boldsymbol{X})|X_k)] &\le \mathbb{E}\mathbb{E}\big[w_{-k|k}(\boldsymbol{X}) \nabla_{-k}^{\mathrm{t}} g(\boldsymbol{X})\, \boldsymbol{\Sigma}_{-k|k} \nabla_{-k} g(\boldsymbol{X})\big|X_k\big] \\ &= \mathbb{E}\big[w_{-k|k}(\boldsymbol{X}) \nabla_{-k}^{\mathrm{t}} g(\boldsymbol{X})\, \boldsymbol{\Sigma}_{-k|k} \nabla_{-k} g(\boldsymbol{X})\big] \\ &= \mathbb{E}\big[w_{-k|k}(\boldsymbol{X})\big(\textstyle\sum_{i=1}^{k-1} \sigma_{i|k}^2 g_i^2(\boldsymbol{X}) + 2\sum_{1\le i<j\le k-1} \sigma_{ij|k} g_i(\boldsymbol{X}) g_j(\boldsymbol{X})\big)\big]. \end{aligned} \tag{3.7}$$

From (3.2), via (3.5) and (3.7), we have that

$$\begin{aligned} \mathrm{Var}\, g(\boldsymbol{X}) \le \mathbb{E}\big[&w(\boldsymbol{X})\sigma_k^2 g_k^2(\boldsymbol{X}) + \textstyle\sum_{i=1}^{k-1}[w(\boldsymbol{X})\sigma_k^2\alpha_k^2 c_{i|k}^2 + w_{-k|k}(\boldsymbol{X})\sigma_{i|k}^2] g_i^2(\boldsymbol{X}) \\ &+2\textstyle\sum_{i=1}^{k-1} w(\boldsymbol{X})\sigma_k^2\alpha_k c_{i|k} g_i(\boldsymbol{X}) g_k(\boldsymbol{X}) \\ &+2\textstyle\sum_{1\le i<j\le k-1}[w(\boldsymbol{X})\sigma_k^2\alpha_k^2 c_{i|k} c_{j|k} + w_{-k|k}(\boldsymbol{X})\sigma_{ij|k}] g_i(\boldsymbol{X}) g_j(\boldsymbol{X})\big]. \end{aligned}$$

After some algebra (see Appendix B) it follows that

$$\mathrm{Var}\, g(\boldsymbol{X}) \le \mathbb{E}\big[w(\boldsymbol{X})\big(\textstyle\sum_{i=1}^{k} \sigma_i^2 g_i^2(\boldsymbol{X}) + 2\sum_{1\le i<j\le k} \sigma_{ij} g_i(\boldsymbol{X}) g_j(\boldsymbol{X})\big)\big]$$

and (3.1) is proved.

Consider the function $g(\boldsymbol{x}) = \rho_0 + \sum_{i=1}^{k} \rho_i x_i$. One can easily see that (3.1) holds as equality. Conversely, assume that (3.1) holds as equality. Then (3.3), (3.4) and (3.6) hold as equalities. From the equality in (3.6), under the inductional assumption, it follows that $g(\boldsymbol{x}) = \varrho_0(x_k) + \sum_{i=1}^{k-1} \varrho_i(x_k) x_i$. From the equality in (3.4) we have that the quantity $g_k(\boldsymbol{x}) + \alpha_k \sum_{i=1}^{k-1} c_{i|k} g_i(\boldsymbol{x})$ is a constant in $x_1, \ldots, x_{k-1}$. Combining the above relations it follows that the quantity $\Delta_k \varrho_0(x_k) + \sum_{i=1}^{k-1}[\Delta_k \varrho_i(x_k)] x_i + \alpha_k \sum_{i=1}^{k-1} c_{i|k} \varrho_i(x_k) = \sum_{i=1}^{k-1}[\Delta_k \varrho_i(x_k)] x_i + h(x_k)$ is a constant in $x_1, \ldots, x_{k-1}$. Therefore, $\Delta_k \varrho_i(x_k) = 0$ for all $i = 1, \ldots, k-1$, that is $\varrho_i(x_k) = \rho_i$, $i = 1, \ldots, k-1$, are constants. Thus, $g(\boldsymbol{x}) = \varrho_0(x_k) + \sum_{i=1}^{k-1} \rho_i x_i$. Finally, from the equality in (3.3) it follows that the quantity $\mathbb{E}(g(\boldsymbol{X})|X_k = x_k)$ is a linear function in x_k. Moreover, $\mathbb{E}(g(\boldsymbol{X})|X_k = x_k) = \mathbb{E}\big(\varrho_0(X_k) + \sum_{i=1}^{k-1} \rho_i X_i | X_k = x_k\big) = \varrho_0(x_k) + \sum_{i=1}^{k-1} \rho_i \mathbb{E}(X_i|X_k = x_k)$. For both cases the quantity $\sum_{i=1}^{k-1} \rho_i \mathbb{E}(X_i|X_k = x_k)$ is a linear function of x_k. Hence, $\varrho_0(x_k)$ is a linear function of x_k, i.e. $\varrho_0(x_k) = \rho_0 + \rho_k x_k$, and the proof is complete. □

A The discrete multivariate covariance identity and some useful properties

Let $\boldsymbol{X}$ a k-dimensional random vector with probability mass function p supported by a "convex" set $C^k \subseteq \mathbb{N}^k$ such that $\mathbf{0} = (0, \ldots, 0)^{\mathrm{t}} \in C^k$ (in the sense the if $\boldsymbol{x} = (x_1, \ldots, x_k) \in C^k$ then $\{0, \ldots, x_1\} \times \cdots \times \{0, \ldots, x_k\} \subseteq C^k$). Assume that the mean $\boldsymbol{\mu}$ and the dispersion matrix $\boldsymbol{\Sigma}$ of $\boldsymbol{X}$ are well defined ($\boldsymbol{\Sigma} > 0$) and consider the vector of linear functions

$$\boldsymbol{q}(\boldsymbol{x}) \equiv \left(q^1(\boldsymbol{x}), \ldots, q^k(\boldsymbol{x})\right)^{\mathrm{t}} := \boldsymbol{\Sigma}^{-1}\boldsymbol{x}. \tag{A.8}$$

Then the $\boldsymbol{w}$-function of $\boldsymbol{X}$ is well defined for every $\boldsymbol{x} \in C^k$ by $\boldsymbol{w}(\boldsymbol{x}) \equiv \left(w^1(\boldsymbol{x}), \ldots, w^k(\boldsymbol{x})\right)$ with

$$w^i(\boldsymbol{x})p(\boldsymbol{x}) = \textstyle\sum_{j=0}^{x_i}[\mu^i - q^i(\boldsymbol{x}_{x_i \mapsto j})]p(\boldsymbol{x}_{x_i \mapsto j}), \tag{A.9}$$

where $\mu^i = \mathbb{E}q^i(\boldsymbol{X})$ and $\boldsymbol{x}_{x_i \mapsto j} = (x_1, \ldots, x_{i-1}, j, x_{i+1}, \ldots, x_k)$ for $i = 1, \ldots, k$ (see Cacoullos and Papathanasiou [11], Papadatos and Papathanasiou [21]).

Cacoullos and Papathanasiou [11] established the identity

$$\mathrm{Cov}[q^i(\boldsymbol{X}), g(\boldsymbol{X})] = \mathbb{E}w^i(\boldsymbol{X})g_i(\boldsymbol{X}), \tag{A.10}$$

provided that $\mathbb{E}|w^i(\boldsymbol{X})g_i(\boldsymbol{X})| < \infty$ and $\mathbb{E}|(q^i(\boldsymbol{X}) - \mu^i)g(\boldsymbol{X})| < \infty$, $i = 1, 2, \ldots, k$.

B Necessary algebra for the proof of Theorem 3.1

Multinomial case:

First, we calculate $c_{i|k} = \sigma^2_{i|k} + \displaystyle\sum_{j=1, j\neq i}^{k-1} \sigma_{ij|k} = \frac{(n-x_k)\pi_i(1-\pi_i-\pi_k)}{(1-\pi_k)^2} + \sum_{j=1, j\neq i}^{k-1} \frac{-(n-x_k)\pi_i\pi_j}{(1-\pi_k)^2} = \frac{(n-x_k)\pi_0\pi_i}{(1-\pi_k)^2}$.

Hence:

$$\begin{aligned}
&w(\boldsymbol{x})\sigma_k^2\alpha_k^2c_{i|k}^2 + w_{-k|k}(\boldsymbol{x})\sigma^2_{i|k} \\
&\quad = \frac{n - \sum_{s=1}^k x_s}{n\pi_0} n\pi_k(1-\pi_k)\frac{(1-\pi_k)^2}{\pi_0^2(n-x_k)^2}\frac{(n-x_k)^2\pi_0^2\pi_i^2}{(1-\pi_k)^4} + \frac{(1-\pi_k)(n-\sum_{s=1}^k x_s)}{(n-x_k)\pi_0}\frac{(n-x_k)\pi_i(1-\pi_i-\pi_k)}{(1-\pi_k)^2} \\
&\quad = \frac{n - \sum_{s=1}^k x_s}{n\pi_0}\left(\frac{n\pi_i^2\pi_k}{1-\pi_k} + \frac{n\pi_i(1-\pi_i-\pi_k)}{1-\pi_k}\right) = w(\boldsymbol{x})n\pi_i\frac{\pi_i\pi_k + 1 - \pi_i - \pi_k}{1-\pi_k} = w(\boldsymbol{x})n\pi_i(1-\pi_i) = w(\boldsymbol{x})\sigma_i^2,
\end{aligned}$$

$$w(\boldsymbol{x})\sigma_k^2\alpha_k c_{i|k} = w(\boldsymbol{x})n\pi_k(1-\pi_k)\frac{-(1-\pi_k)}{\pi_0(n-x_k)}\frac{(n-x_k)\pi_0\pi_i}{(1-\pi_k)^2} = w(\boldsymbol{x})(-n\pi_i\pi_k) = w(\boldsymbol{x})\sigma_{ik},$$

$$\begin{aligned}
&w(\boldsymbol{x})\sigma_k^2\alpha_k^2c_{i|k}c_{j|k} + w_{-k|k}(\boldsymbol{x})\sigma_{ij|k} \\
&\quad = \frac{n - \sum_{s=1}^k x_s}{n\pi_0} n\pi_k(1-\pi_k)\frac{(1-\pi_k)^2}{\pi_0^2(n-x_k)^2}\frac{(n-x_k)^2\pi_0^2\pi_i\pi_j}{(1-\pi_k)^4}\frac{(1-\pi_k)(n-\sum_{s=1}^k x_s)}{(n-x_k)\pi_0}\frac{-(n-x_k)\pi_i\pi_j}{(1-\pi_k)^2} \\
&\quad = \frac{n - \sum_{s=1}^k x_s}{n\pi_0}\left(\frac{n\pi_i\pi_j\pi_k}{1-\pi_k} - \frac{n\pi_i\pi_j}{1-\pi_k}\right) = w(\boldsymbol{x})(-n\pi_i\pi_j) = w(\boldsymbol{x})\sigma_{ij}.
\end{aligned}$$

Negative Multinomial case:

We calculate $c_{i|k} = \sigma^2_{i|k} + \displaystyle\sum_{j=1, j\neq i}^{k-1} \sigma_{ij|k} = \frac{(r+x_k)\theta_i(\theta_0+\theta_i+\theta_k)}{(\theta_0+\theta_k)^2} + \sum_{j=1, j\neq i}^{k-1} \frac{(r+x_k)\theta_i\theta_j}{(\theta_0+\theta_k)^2} = \frac{(r+x_k)\theta_i}{(\theta_0+\theta_k)^2}$. Hence:

$$\begin{aligned}
&w(\boldsymbol{x})\sigma_k^2\alpha_k^2c_{i|k}^2 + w_{-k|k}(\boldsymbol{x})\sigma^2_{i|k} \\
&\quad = \frac{\theta_0(r + \sum_{s=1}^k x_s)}{r}\frac{r\theta_k(\theta_0+\theta_k)}{\theta_0^2}\frac{(\theta_0+\theta_k)^2}{(r+x_k)^2}\frac{(r+x_k)^2\theta_i^2}{(\theta_0+\theta_k)^4} + \frac{(\theta_0+\theta_k)(r+\sum_{s=1}^k x_s)}{(r+x_k)}\frac{(r+x_k)\theta_i(\theta_0+\theta_i+\theta_k)}{(\theta_0+\theta_k)^2} \\
&\quad = \frac{\theta_0(r + \sum_{s=1}^k x_s)}{r}\left(\frac{r\theta_i^2\theta_k}{\theta_0^2(\theta_0+\theta_k)} + \frac{r\theta_i(\theta_0+\theta_i+\theta_k)}{\theta_0(\theta_0+\theta_k)}\right) = w(\boldsymbol{x})\frac{r\theta_i(\theta_0+\theta_i)}{\theta_0^2} = w(\boldsymbol{x})\sigma_i^2,
\end{aligned}$$

$$w(\boldsymbol{x})\sigma_k^2\alpha_k c_{i|k} = w(\boldsymbol{x})\frac{r\theta_k(\theta_0+\theta_k)}{\theta_0^2}\frac{\theta_0+\theta_k}{r+x_k}\frac{(r+x_k)\theta_i}{(\theta_0+\theta_k)^2} = w(\boldsymbol{x})\frac{r\theta_i\theta_k}{\theta_0^2} = w(\boldsymbol{x})\sigma_{ik},$$

$$
\begin{aligned}
&w(\boldsymbol{x})\sigma_k^2\alpha_k^2 c_{i|k}c_{j|k} + w_{-k|k}(\boldsymbol{x})\sigma_{ij|k}\\
&\quad= \frac{\theta_0(r+\sum_{s=1}^k x_s)}{r}\frac{r\theta_k(\theta_0+\theta_k)}{\theta_0^2}\frac{(\theta_0+\theta_k)^2}{(r+x_k)^2}\frac{(r+x_k)^2\theta_i\theta_j}{(\theta_0+\theta_k)^4} + \frac{(\theta_0+\theta_k)(r+\sum_{s=1}^k x_s)}{(r+x_k)}\frac{(r+x_k)\theta_i\theta_j}{(\theta_0+\theta_k)^2}\\
&\quad= \frac{\theta_0(r+\sum_{s=1}^k x_s)}{r}\Big(\frac{r\theta_i\theta_j\theta_k}{\theta_0^2(\theta_0+\theta_k)} + \frac{r\theta_i\theta_j}{\theta_0(\theta_0+\theta_k)}\Big) = w(\boldsymbol{x})\frac{r\theta_i(\theta_0+\theta_i)}{\theta_0^2} = w(\boldsymbol{x})\sigma_{ij}.
\end{aligned}
$$

References

[1] Afendras, G. (2011). Universal variance bounds for the Pearson family. *Submitted.* arXiv: math.PR/1110.0090

[2] Afendras, G. and Papadatos, N. (2011). On matrix variance inequalities, *J. Statist. Plann. Inference* **141** 3628–3631.

[3] Afendras, G. and Papadatos, N. (2011). Strengthened Chernoff-type variance bounds. *Submitted.* arXiv: math.PR/1107.1754

[4] Afendras, G., Papadatos, N. and Papathanasiou, V. (2007). The discrete Mohr and Noll Inequality with applications to variance bounds. *Sankhyā*, **69**, 162–189.

[5] Afendras, G., Papadatos, N. and Papathanasiou, V. (2011). An extended Stein-type covariance identity for the Pearson family, with applications to lower variance bounds. *Bernoulli*, 17(2), 507–529.

[6] Arnold, B.C. and Brockett, P.L. (1988). Variance bounds using a theorem of Polya. *Statist. Probab. Lett.* **6** 321-326.

[7] Brascamp, H.J. and Lieb, E.H. (1976). On extensions of the Brunn–Minkowski and Prékopa–Leindler Theorems, including inequalities for log concave functions, and with application to the diffusion equation. *J. Functional Analysis.* **22** 366–389.

[8] Cacoullos, T. (1982). On upper and lower bounds for the variance of a function of a random variable. *Ann. Probab.* **10** 799–809.

[9] Cacoullos, T. and Papathanasiou, V. (1985). On upper bounds for the variance of function of random variables. *Statist. Probab. Lett.* **3** 175–184.

[10] Cacoullos, T. and Papathanasiou, V. (1989). Characterizations of distributions by variance bounds. *Statist. Probab. Lett.* **7** 351–356.

[11] Cacoullos, T. and Papathanasiou, V. (1992). Lower Variance Bounds and a New Proof of the Central Limit Theorem. *J. Multivariate Anal.* **43** 173–184.

[12] Chang, W.-Y. and Richards, D.ST.P. (1999). Variance Inequalities for Functions of Multivariate Random Variables. *Advances in stochastic inequalities (Atlanta, GA, 1997), Contemp. Math.* **234** Amer. Math. Soc., Providence, RI, 43–67.

[13] Chen, H.Y.L. (1982). An Inequality for the Multivariate Normal Distribution. *J. Multivariate Analysis* **12** 306–315.

[14] Chernoff, H. (1981). A note on inequality involving the normal distribution. *Ann. Probab.* **9** 533–535.

[15] Houdré, C. (1995). Some applications of covariance identities and inequalities to functions of multivariate normal variables. *J. Amer. Statist. Assoc.* **90** 965–968.

[16] Houdré, C. and Kagan, A. (1995). Variance inequalities for functions of Gaussian variables. *J. Theoret. Probab.* **8** 23–30.

[17] Houdré, C. and Pérez-Abreu, V. (1995). Covariance identities and inequalities for functionals on Wiener and Poisson spaces. *Ann. Probab.* **23** 400–419.

[18] Johnson, R.W. (1993). A note on variance bounds for a function of a Pearson variate. *Statist. Decisions* **11** 273–278.

[19] Klaassen, C.A.J. (1985). On an inequality of Chernoff. *Ann. Probab.* **3** 966–974.

[20] Nash, J. (1958). Continuity of solutions of parabolic and elliptic equations. *Ame. J. Math.* **80** 931–954.

[21] Papadatos, N. and Papathanasiou, V. (1998). Variational Inequalities for Arbitrary Multivariate Distributions. *J. Multivariate Anal.* **67** 154–168.

[22] Papathanasiou, V. (1988). Variance bounds by a generalization of the Cauchy-Schwarz inequality. *Statist. Probab. Lett.* **7** 29–33.

[23] Prakasa Rao, B.L.S. (2006). Matrix variance inequalities for multivariate distributions. *Statistical Methodology* **3** 416–430.
[24] Wei, Z. and Zhang, X. (2009). Covariance matrix inequalities for functions of Beta random variables. *Statist. Probab. Lett.* **79** 873–879.